8-21

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Land Cover Boundary Extraction in Rural Aerial Videos

Babak Majidi
Faculty of Engineering and Industrial Sciences,
Swinburne University of Technology,
Hawthorn, Victoria, Australia
bmajidi@swin.edu.au

Alireza Bab-Hadiashar
Faculty of Engineering and Industrial Sciences,
Swinburne University of Technology,
Hawthorn, Victoria, Australia
abab-hadiashar@swin.edu.au

Abstract

In this paper a new approach to finding and tracking various land cover boundaries such as rivers, agricultural fields, channels and roads for use in visual navigation system of an unmanned aerial vehicle is presented. We use a combination of statistical estimation and optimization techniques for extraction of dominant boundaries in noisy aerial images. A set of perceptual grouping restrictions is used to connect the acquired piecewise boundaries and to find the heading direction of the main boundary. The results are further refined by applying a set of texture and colour cues and eliminating any false hypothesis. Our results show our method outperforms single feature object tracking algorithms in this application.

1. Introduction

From the second half of the twentieth century scientist began to appreciate the importance of monitoring the earth conditions. Monitoring information is important as they can reveal problems associated with the environment, such as decline in biodiversity, global warming, soil erosion, deforestation and impact of various pollutants. Importance of such information arises from the fact that it can be utilized to find appropriate remedies for those problems. To deal with such issues, there is a substantial need to identify and understand the rate, source and result of natural changes from both national and local perspective in a timely and accurate manner. Collecting such a timely and accurate picture of various environmental parameters contributes to damage minimization strategies and assists environmental preservation activities.

One of the promising tools for automation of environmental monitoring is the use of unmanned aerial vehicles (UAV). A simple UAV can be used as a robotic ranger for monitoring natural resources over vast areas. Such systems can fly with altitude of 5m to 100m and upon detection of desired environmental features, perform appropriate pre-programmed manoeuvres depending on the type of application they are used for. These manoeuvres can either be tracking a desired feature, increasing or decreasing the altitude of the plane to view features from a vantage point, or a simple path planning to scan the area for detection of similar features. The visual navigation system of this UAV must be capable of interpretation of mostly rural natural scenes from low quality images captured by on board cameras. It's critical for the vision system of such a UAV to be able to follow boundaries of various land covers.

Currently there is a large literature on object detection in aerial images. However most of the existing algorithms for remote sensing are designed for satellite and high-altitude airborne imaging systems, and will not be appropriate to UAV applications and so most of those algorithms need to be reconsidered for the proposed task. Also current UAV visual navigation projects [1] are mostly focused on road following in an urban environment. Since our goal is focused on rural areas, most of the assumptions of those algorithms are not appropriate for our application. Natural land covers are usually don't have smooth boundaries. Moreover, their edges are rarely continuous and are often partially hidden (from the above) by other land covers such and trees and bushes. They don't have simple geometrical shapes and cannot be described by models that are used for manmade structures.

In particular, the general assumption for road detection in aerial images which states that a road consists of two parallel lines is not applicable to most of our desired entities. We therefore propose to use a combination of edge and texture based cues in order to follow natural boundaries. In our method, first we find dominant land cover boundaries using a combination of edge detection and a robust estimation technique and then use colour and texture information to eliminate any possible false hypotheses and extract the boundary of the main natural land cover that UAV is likely to be following.

In the remainder of this paper, we first discuss the extraction of partial boundaries and describe how we can connect these partial boundaries with perceptual grouping. Then generation of texture cues and object detection using combination of these methods is discussed in section 4. Section 5 shows the results of our experiments on a number of aerial videos taken by an onboard camera of a small UAV. Section 6 concludes the paper.

2. Boundary Extraction Using a Robust Estimation Technique

To find the dominant boundaries, we have developed an elaborate scheme that combines local edge detection with global robust line fitting strategies. The detail description of our approach is outside the scope of this paper but we briefly outline the main steps and show its results. The first step in our scheme for extracting dominant boundaries is to determine an approximate orientation of the UAV. We can find this orientation either by using the heading angle of UAV or by calculating the cross correlation between two successive frames (the latter is used here). Having the direction of motion, we filter the image by a low pass filter (combination of median and Gaussian convolutions) and apply a directional edge detection algorithm (a variation of the Sobel mask). We then employ an edge linking algorithms [2] to combine small edges to generate longer ones and apply the Modified Selective Statistical Estimator (MSSE) [3] to find the global lines that best fit the edge segments. MSSE is a very straight forward technique and since we know a number of priories including the size of images and approximate desired direction of objects, we have been able to implement the robust estimation step in a novel way (using the Downhill Simplex method [4]) which significantly increases the speed and accuracy of the estimation process. Moreover, the number of edge segments supporting the global lines is very limited in practice and therefore the finite sample bias of the chosen estimator is of concern. Again, MSSE is deemed suitable as it has the least finite sample bias amongst the currently available robust estimators [5]. Due to space limitation, the detail of our new implementation of the MSSE is not included in this paper.

Figure 1 shows the result of this stage of our algorithm applied to an aerial image taken by our UAV following a river. This figure shows that the proposed scheme has been able to identify the main edge components. In the next step, we somehow need to process these segments and find the ones that truly represent object boundaries. To achieve this, we will refine and implement a set of perceptual grouping algorithms to connect line segments and generate the object boundaries.

Figure 1: Our original frame of aerial video and extracted lines by the MSSE.

3. Perceptual Grouping of Extracted Lines

A common approach to for connecting line segments is to use perceptual grouping concept. Manmade objects in aerial images are usually modelled by a combination of edge based features: short and long straight lines, short and long curved lines, multiple line junctions, L shape junctions, U shape junctions, parallel lines and closed polygons [6]. Such well-defined entities can be generated by classic perceptual grouping algorithms [7]. In contrast to manmade (urban) objects, rural land covers can take variety of shapes and usually don't consist of sharp edges and straight lines. For instance, the boundary of rivers are usually modelled by piecewise continuous curved lines while agricultural fields can be reconstructed using a combination of piece wise straight and curved lines. Small patches in natural environment are usually modelled by deformed polygons.

The first issue arise in this context relates to the fact that the line extraction process described earlier usually produces small superfluous lines. Such lines do not have spatial support and are not significant in describing the boundaries of objects and should therefore be eliminated. The elimination process involves a simple thresholding of the length of spatial support.

Having obtained a set of relatively long linear segments representing the most dominant object boundaries in an image, the problem is that these lines are not connected while some semantically meaningless line segments also appear (false hypothesis). Therefore, the main task in the next steps is to link representative line segments and to eliminate false hypotheses. To resolve this issue, we modify the perceptual grouping rules of detecting manmade objects to a new set specially designed to deal with the lack of prior knowledge about the shape of natural objects in rural areas. Application of these rules would result in closing of gaps and elimination of false hypotheses.

The first step in perceptual grouping of piecewise boundaries is to close gaps between line segments by applying the proximity function. Approximately collinear edges with endpoints closer than a predefined threshold are connected to each other at this stage. Due to the irregularities of natural boundaries, the two recovered part of a line do not exactly match each other and therefore the proximity concept needs to be extended. Therefore, if two parallel lines satisfy the proximity function and are closer than a threshold we assume that they are part of the same line.

The common perceptual grouping rules for object detection are mostly developed for extraction of manmade objects and are very much dependent upon the precise geometry and engineered symmetry of such objects. For instance, rooftop detection algorithms assume that all rooftops are rectangles while road detection algorithms rely on the fact that road boundaries are parallel. For natural objects these assumptions are often false and so need to be changed to reflect the broad conditions of such objects.

Categorization into predefined line junctions is the common approach to connecting line segments. A junction point is the intersection of two lines obtained by extending the endpoints which fall in a common region. Various kinds of junctions are described in literature [6, 8]. For our special purpose we selected two basic junctions (L and U junctions) and generalise their description for our application, named them: V shape and trapezoidal shape junctions.

A V junction is either represents a corner of a planar entity (appearance of specific vegetations) or a sharp change in the boundary of an elongated object (a turn in a river). A trapezoidal junction is formed by the alignment of two V junctions and are generated by sharp bends in boundaries (a river turning back) or smaller natural features that has not yet formed a close polygon. Two types of trapezoidal junction are common: two V junctions with or without a common line. To find the instances of these line junctions, a small square neighbourhood centred at the intersection of main lines is searched.

It is generally straightforward to connect different pieces of a boundary by the aforementioned grouping rules. However, there are cases in which the intersection of a junction is partially covered by an object or its shadow. When a boundary is broken (covered) along its straight parts, there is usually enough support to fit a line to its remaining parts and recover the missing part. Serious difficulties arise when a junction with a significant change of direction is covered (missing). For example, in Figure 1 the lower boundary of the river at a turning point is partially covered by some shadows (in the middle of this image) and tree foliages. Although it would be possible in some case to use shadow elimination algorithms, it is generally not that useful, as such remedies would not be effective for other types of covering artefacts. The use of prescribed grouping rules to follow the boundary of the river shown in Figure 1 will completely fail due to existence of shadow patches at the important junctions (river twists). To solve this problem, we propose to use a hybrid colour-texture segmentation algorithm to rectify the problem.

4. Boundary Extraction: Hybrid Approach

As we mentioned earlier, boundaries of natural objects viewed from above are often broken. It's common to have long patches of trees covering boundaries of roads, agricultural fields and rivers which make it almost impossible to detect the actual orientation of a boundary after such gaps. A boundary may or may not continue in the same orientation after a gap and in the worst case scenario, there may not be even a hint of an orientation change in either sides of the gap. In such a case, a possible solution would be to use other properties of the tracked land cover.

In our approach, we use a simple colour and texture segmentation algorithm to generate rough map of various land covers in an image. By combining this map with piecewise boundary map that was generated in previous section, we would be able to determine the exact direction of the tracked land cover.

Similar to the case of perceptual grouping rules, segmentation of natural objects is also much harder than man made ones. The basic reason for this is the nature of objects being imaged: these objects do not have well defined boundaries. Moreover, inadequate lighting diminishes the intensity (colour or grey alike) contrast of an object and its background, making it difficult for a segmentation algorithm to generate an accurate boundary.

In the first stage of our segmentation algorithm, we manually create a database of commonly plausible land covers of interest and categorize them in a set of predefined classes named: Tree, grass, cloud, sky, road and river. Although we are interested in tracking land covers, some camera angles and specific manoeuvres of the capturing UAV will lead to appearance of sky and cloud in the captured video.

We then define the dominant characteristics of every land cover class. Such characteristics should be easily calculable and differentiating across various classes. For example, if we have one hundred instances of class 'tree', then we can derive a number of features from these instances that define a tree class. In future if a tree appears in a new image, its class can easily be identified as a tree using a feature matching technique.

Having developed a set of characteristic features for our land cover classes, their suitability for classification of encountered land covers, seen from an aerial platform, must be examined. The examination involves using a training data set for tuning the method to recognize natural objects of existing samples and recording its success in recognizing new instances of those land covers in other similar videos.

In our implementation, a database of 1000 aerial video frames taken from various platforms is used for training. These frames consist of variety of semi-urban (having some manmade part) and rural scenes. A major problem with rural images is that, unlike most manmade objects, natural objects don't have uniform colour and texture characteristics. Roads and rooftops usually have relatively uniform textures but even in a single video, one usually finds many different classes of trees and bushes with various colour and texture characteristics. In this study, we have mainly used images that are taken in good conditions minimizing the effect of noise. However, as our goal is to apply our techniques to real world problems, some images which captured in less than ideal conditions have also been included. Even though such images represent a more realistic view of what kind of images an automated object recognition system must cope with, they add to the complexity of the classification task.

As it was mentioned earlier, our classification uses six different classes (tree, grass, river, cloud, sky and road). Some of these classes are fairly homogeneous in their texture, such as river, sky, and road, whereas others are more varied, such as tree, grass and cloud. Also, the appearance of common vegetations usually changes significantly under different lighting conditions and so increasing the variability in samples. Clouds are by nature varied and thus their samples do not form a very homogeneous class.

In our algorithm, classification of natural objects is performed by first separating vegetation from the rest of the data as vegetation data is usually overlap with other land covers (it is common to see vegetation on the side roads, rivers, etc). The identification of vegetation parts is performed by processing the amount of green colour using a simple colour histogram thresholding technique. Since it is relatively straightforward to segment vegetation from other land covers, the two-stage implementation of the proposed classification framework significantly improves the classification outcomes.

Once the vegetation parts are identified, two different classifiers can be trained to categorize two vegetation classes by one and remaining four natural object classes by the other. Figure 2 shows result of our colour and texture segmentation applied to the image shown in Figure 1.

As was shown in section 3, the perceptual grouping of boundary pieces alone is not sufficient to overcome the existing gaps along the boundaries of elongated objects particularly at places where such objects turn and twist. We proposed to improve the situation by using a combination of perceptual grouping and cues provided by colour and texture segmentation.

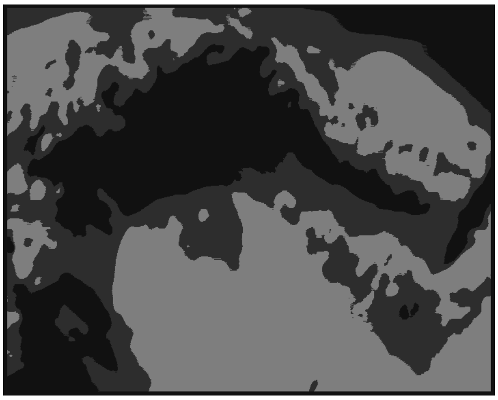

Figure 2: Results of our colour and texture segmentation algorithm applied to the image shown in Figure 1.

To combine those approaches, we propose to use a very simple case based reasoning system to eliminate possible false hypotheses and extract the actual boundary of the desired land cover. The reasoning system excludes the impossible combinations of various cues and provides a direction by which the algorithm can join the appropriate line segments. Figure 3 shows how the colour-texture segmentation results help the algorithm to follow the river as it turns. It is important to note that the river boundary at its turning point is completely covered by shadows and no meaningful line segments are available.

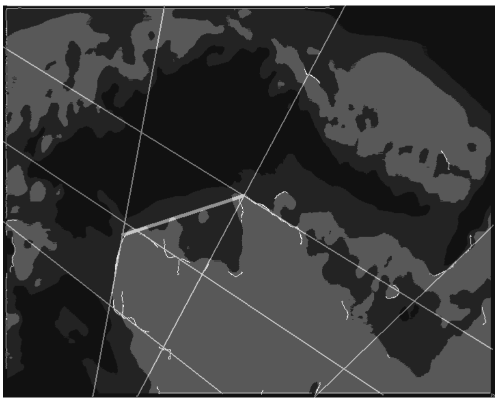

Figure 3: Finding the right direction of boundary combining perceptual grouping and colour-texture segmentation results.

5. Experimental Results

Boundary detection of natural objects in aerial images is a challenging task and a comprehensive solution remains elusive. To verify the performance of the proposed technique, we have applied our algorithm to 100 aerial images of several videos captured by flying platforms. Many of those images are taken by a Canon MV400 camera that is carried by our drone flying over a river. In many cases, we have been able to follow the desired boundary using only one set of parameters. However, tuning a right set of parameters to maximise the performance of the boundary detection algorithm remains an issue. There are also instances of very complicated scenarios that our algorithm detects and follows a false boundary. We think that a more sophisticated case based reasoning system is required to handle such cases.

6. Conclusions

In this paper, we presented a procedure for automatic boundary extraction of various land covers in videos captured by a UAV flying at low altitudes. Our focus has been on the mapping of elongated objects and we have studied a number of important issues that normally arise in tracking such objects. A hybrid solution combining the results of a robust line detection and a colour-texture segmentation algorithms is developed. The method is capable of detecting various land cover boundaries in rural aerial video and has been tested using 100 aerial images of various land covers. The results show that the hybrid approach has been able to solve a number of important problems particularly with regards to sudden change of direction in places covered by other artefacts.

7. References

[1] E. Frew, T. McGee, Z. Kim, X. Xiao, S. Jackson, M. Morimoto, S. Rathinam, J. Padial, and R. Sengupta, "Vision-based Road Following Using a Small Autonomous Aircraft", Proc. IEEE Aerospace Conference, 2004.

[2] P. D. Kovesi., MATLAB and Octave Functions for Computer Vision and Image Processing. School of Computer Science & Software Engineering, The University of Western Australia.

[3] Bab-Hadiashar A., Suter D., "Robust Segmentation of Visual Data Using Ranked Unbiased Scale Estimator", International Journal of Information, Education and Research in Robotics and Artificial Intelligence, ROBOTICA, volume 17, pp. 649-660, 1999.

[4] Nelder, J.A., and Mead, R., "A simplex method for function minimization", The Computer Journal, 7, 308-313, 1964.

[5] R. Hoseinnezhad, A. Bab-Hadiashar and D. Suter, "Finite Sample Bias of Robust Scale Estimators in Computer Vision Problems," International Symposium on Visual Computing (ISVC'06), pp. 445-454, USA, 2006.

[6] Qasim Iqbal, J. K. Aggarwal, "Retrieval By Classification of Images Containing Large Manmade Objects Using Perceptual Grouping", Pattern recognition, vol. 35, No. 7, pp. 1463-1479, 2002.

[7] Sarkar, S. Boyer, K.L., "Perceptual organization in computer vision: a review and a proposalfor a classificatory structure", IEEE Transactions on Systems, Man and Cybernetics, Volume: 23, Issue: 2, pp. 382-399, Mar/Apr 1993.

[8] Xiaofen Zheng Qigang Gao, "Detection of perceptual junctions by curve partitioning and grouping", Proceedings of First Canadian Conference on Computer and Robot Vision, pp. 347- 353, May 17-19, 2004.